SORT 45 (2) July-December 2021, 1-18 DOI:

Median bilinear models in the presence of extreme values

Miguel Santolino*

Abstract

Bilinear regression models involving a nonlinear interaction term are applied in many fields (e.g., Goodman's RC model, Lee-Carter mortality model or CAPM financial model). In many of these contexts data often exhibit extreme values. We propose the use of bilinear models to estimate the median of the conditional distribution in the presence of extreme values. The aim of this paper is to provide alternative methods to estimate median bilinear models. A calibration strategy based on an iterative estimation process of a sequence of median linear regression is developed. Mean and median bilinear models are compared in two applications with extreme observations. The first application deals with simulated data. The second application refers to Spanish mortality data involving years with atypical high mortality (Spanish flu, civil war and HIV/AIDS). The performance of the median bilinear model was superior to that of the mean bilinear model. Median bilinear models may be a good alternative to mean bilinear models in the presence of extreme values when the centre of the conditional distribution is of interest.

MSC: 62H12, 62H17, 62JO2, 62L12.

Keywords: Outliers, quantile regression, single factor models, nonlinear, multiplicative.

1. Introduction

In regression analysis the effect of the interaction between two explanatory variables on the dependent variable is often of great interest. Two-way analysis of variance (ANOVA) models have been widely applied in linear regression analysis when a measurement dependent variable is regressed on two categorical independent variables, and the aim is to assess the main effect of the two nominal variables but also the interaction effect between them (Yates and Cochran, 1938). Two-way ANOVA models are linear

* Riskcenter-IREA, Dept. Econometrics, University of Barcelona, Spain. E-mail: msantolino@ub.edu
Received: February 2021
Accepted: November 2021

regression models where the joint interaction effect is included as an additional regressor. So, linear regression techniques may be directly applied to estimate parameters, such as least squares or maximum likelihood methods.

A more flexible modelling in two-way tables are the regression models in which the multiplicative interaction structure is specified as a nonlinear term. These models are usually named bilinear models (Gabriel, 1978), although other names are often in use for these models, such as biadditive models (Denis and Pázman, 1999) or additive main effects and multiplicative interaction (AMMI) models (Van Eeuwijk, 1992, 1995). The unknown parameters of bilinear models may be also estimated by least squares or maximum likelihood. Least squares estimators of the nonlinear term are derived using singular value decomposition of the matrix of residuals (Gabriel, 1978; Lee and Carter, 1992). Maximum likelihood estimators may be obtained by an iterative process (Goodman, 1979, 1981).

Bilinear regression models involving multiplicatively structured interactions are widely applied. Many models used in social sciences fits to this setting, including the row-column association model for two-way tables (Goodman, 1979, 1981), the uniform difference (UNIDIFF) or layer effect model for three-way tables (Erikson and Goldthorpe, 1992; Xie, 1992), generalized additive main effects and multiplicative interaction effects (GAMMI) models for crop yields (Van Eeuwijk, 1992, 1995), the one-dimensional Rasch-type model for binary responses (Turner, Firth and Kosmidis, 2013) or the stereotype regression model for ordered multinomial data (Anderson, 1984). In time series analysis, statistical factor models can be understood as multiplicative interaction models (Croux et al., 2003). Factor models are widely applied in finance for calculating the investment risk in asset pricing theory, such as the capital asset pricing model (CAPM) model or the Fama-French model (Black, Jensen and Scholes, 1972). In demography and actuarial science, factor models are used to predict the future mortality. In fact, most of mortality projections models, such as Lee-Carter and Renshaw-Haberman models can be understood as multiplicative interaction models (Lee and Carter, 1992; Renshaw and Haberman, 2006; Macias and Santolino, 2018; Moyano-Silva et al., 2020).

In many of these contexts data often show extreme values. When the centre of the conditional distribution is of interest, a common practice is to consider extreme values as outliers and remove them from the dataset prior to estimation. Formally, an outlier is a data point that deviates so far from the other observations because it was generated by a totally different mechanism or simply by error (Hawkins, 1980; Justel, Peña and Tay, 2001). Deleting outliers is important because those values can increase error variance and influence estimates. However, this strategy should be taken very cautiously when data points are extreme values but not outliers. Extreme values are events that might happen, so we should be very cautious before deleting these values from datasets.

A different approach is here followed. It is well known that the median is a robust measure of central tendency. Median bilinear models may be a good alternative to mean bilinear models in the presence of extreme values (Gabriel and Odoroff, 1984). In this article we propose the use of the bilinear regression setting to model the median of the conditional distribution as a nonlinear function of predictors. The aim of this article is twofold: 1) to show how the parameters of the median bilinear model can be estimated and, 2) to compare the performance of the conditional median bilinear regression and the conditional mean bilinear regression in the presence of extreme values.

The main contribution of the paper is to review alternative methods to estimate median bilinear models. Bilinear models are nonlinear regressions. The techniques available for estimating nonlinear regression models for the conditional median are not as well developed as those for the conditional mean estimation. Koenker and Park (1996) proposed to calibrate median nonlinear regression models by means of the linearization of the objetive function. Here we propose an alternative calibration approach based on an iterative estimation process of a sequence of median linear regressions. This second alternative is novel. It was first used by Moyano-Silva, Pérez-Marín and Santolino (2020) to estimate the Lee-Carter stochastic mortality model. We here generalize this strategy to estimate median bilinear models with two main factors. To solve the underlying linear optimization problems, we use interior point methods (Koenker and Park, 1996; Portnoy and Koenker, 1997) and the maximum likelihood approach (Sánchez, Labros and Labra, 2013). This paper focuses on the evaluation of goodness-of-fit of mean and median bilinear models in presence of extreme values. However, bootstrapping techniques can be used to estimate standard errors when inference on coefficient estimates is of interest (Buchinsky, 1995).

Two applications are illustrated for the comparison of the median and mean bilinear models. The first application is based on simulated data. In the second application real Spanish mortality data are used to estimate the median and mean (log)bilinear stochastic mortality models. In both applications, bilinear models are calibrated using the whole sample. The performance of the fitted models is then evaluated computing a series of goodness-of-fit measures for the whole sample and when extreme values are removed.

The article is structured as follows. Section 2 introduces the mean and median bilinear regression models. Section 3 shows the parameter estimation methods of the mean bilinear regression model. Section 4 describes the calibration strategies of the median bilinear regression model. The two applications are illustrated in Section 5. Main conclusions are summarized in Section 6.

2. Bilinear regression model

Let Y be a continuous random variable with finite expectation and cumulative distribution function F_Y defined by $F_Y(y) = P(Y \leq y)$. The inverse function of F_Y is known as *quantile function*, Q. The quantile of order α is defined as $Q_\alpha(Y) = F_Y^{-1}(\alpha) = \inf\{y \mid F_Y(y) \geq \alpha\}$ where $\alpha \in (0,1)$. The quantile is a left-continuous increasing function. If F_Y is continuous and strictly increasing, the mathematical expectation can be represented as $\mathbb{E}(Y) = \int_0^1 Q_{1-u}(Y)du$. The median is the quantile of order 0.5.

Let consider two categorical variables. The first factor has I levels $(i \in \{1,\ldots,I\})$ and the second has J levels $(j \in \{1,\ldots,J\})$. The sample size is N such as $N = I \cdot J$.

Let y_{ij} be the random variable conditional on the levels i and j. The bilinear regression model in two-way tables is defined as:

$$y_{ij} = a_i + b_j + c_i \cdot d_j + \varepsilon_{ij} \tag{1}$$

where a_i is the main effect of the level i and b_j is the main effect of the level j. The coefficients of the nonlinear term are c_i and d_j capturing the interaction effect of the two levels. Finally, ε_{ij} is the error random variable. Note that infinite solutions exist in (1). For any scalars z, u and v, the following transformations $\{\tilde{a}_i, \tilde{b}_j, \tilde{c}_i, \tilde{d}_j\} = \{a_i - z \cdot c_i, b_j - d_j \cdot v - z \cdot v, \dfrac{c_i + v}{u}, u \cdot (d_j + z)\}$ give unaltered outcome values. To overcome the lack of identifiability and to help in the interpretation, the following two constraints are often set: $\sum_i c_i = 1$ and $\sum_j d_j = 0$.

In the case of independent and identically zero-mean distributed random errors, the conditional expected value of y_{ij} may be expressed as

$$\mathbb{E}(y_{ij}) = a_i + b_j + c_i \cdot d_j \tag{2}$$

Analogously, in case of independent and identically zero-median distributed random errors, the median of y_{ij} may be expressed as (Bassett and Koenker, 1978),

$$Q_{0.5}(y_{ij}) = a_i + b_j + c_i \cdot d_j \tag{3}$$

Sections 3 and 4 are devoted to estimate the vectors of coefficients, $a = (a_1, \ldots, a_I)$, $b = (b_1, \ldots, b_J)$, $c = (c_1, \ldots, c_I)$ and $d = (d_1, \ldots, d_J)$ in (2) and (3), respectively.

3. Mean bilinear model: calibration

Two widely used techniques to estimate the parameters of (2) are the least squares and the maximum likelihood methods.

Least squared errors

The expectation is the value that minimizes the sum of squared deviations. One strategy for estimating the parameters is to minimize the sum of squared errors, as follows:

$$\min_{\theta \in \mathbb{R}^{2\cdot(I+J)}} \sum_{i,j} (y_{ij} - a_i - b_j - c_i \cdot d_j)^2 \tag{4}$$

where θ is the set of parameters to estimate, $\theta = (a,b,c,d)$. Coefficients in (4) cannot be directly estimated by ordinary least squares because the right-hand side of equation (2) is not linear with the parameters. To estimate the coefficients, Gabriel (1978) proposed to fit the bilinear models in a two-stage process: (1) fit the linear part of the model, then take residuals, and (2) fit the bilinear part to the residuals.

Stage 1 uses linear least squares to solve the least squares problem. The resulting vectors $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_I)$ and $\hat{b} = (\hat{b}_1, \ldots, \hat{b}_J)$ are then introduced into the joint fitting problem. The $(I \times J)$-matrix A, where the (i,j)-element is $a_{ij} = y_{ij} - \hat{a}_i - \hat{b}_j$, is decomposed by singular value decomposition, $\text{svd}(A) = U \Lambda V^{\intercal}$. The vector of estimates $\hat{c} = (\hat{c}_1, \ldots, \hat{c}_I)$ is the first column of U, $\hat{c} = (u_{1,1}, \ldots, u_{I,1})$, and the vector of estimates $\hat{d} = (\hat{d}_1, \ldots, \hat{d}_J)$ is the first column of V multiplied by the first eigenvalue $\lambda_{1,1}$, $\hat{d} = \lambda_{1,1} \cdot (v_{1,1}, \ldots, v_{J,1})$.

Maximum likelihood

Goodman (1979) proposed to use a iterative method for estimating bilinear models by maximum likelihood. Suppose the log-likelihood function is given by $l(\theta) = \sum_{i,j} log\,(f(y_{ij}))$, where f is the density function of y_{ij}. The function l may be maximized by an iterative process in which the elementary newton method is applied for the score functions of each set of parameters. In the mean bilinear model we have three sets of parameters. Denote the vector of initial values $\hat{\theta}^0 = (\hat{a}^0, \hat{b}^0, \hat{c}^0, \hat{d}^0)$ and $l^0 = l(\hat{\theta}^0)$. In the iteration step v, parameters are updated as follows:

1. Given $\hat{\theta}^v$, $\hat{a}^{v+1} = \hat{a}^v - \frac{\partial l^v / \partial a}{\partial^2 l^v / \partial a^2}$, $\hat{b}^{v+1} = \hat{b}^v - \frac{\partial l^v / \partial b}{\partial^2 l^v / \partial b^2}$, $\hat{c}^{v+1} = \hat{c}^v$ and $\hat{d}^{v+1} = \hat{d}^v$.
2. Given $\hat{\theta}^{v+1}$, $\hat{c}^{v+2} = \hat{c}^{v+1} - \frac{\partial l^{v+1} / \partial c}{\partial^2 l^{v+1} / \partial c^2}$ and $\hat{a}^{v+2} = \hat{a}^{v+1}$, $\hat{b}^{v+2} = \hat{b}^{v+1}$ and $\hat{d}^{v+2} = \hat{d}^{v+1}$.
3. Given $\hat{\theta}^{v+2}$, $\hat{d}^{v+3} = \hat{d}^{v+2} - \frac{\partial l^{v+2} / \partial d}{\partial^2 l^{v+2} / \partial d^2}$, $\hat{a}^{v+3} = \hat{a}^{v+2}$, $\hat{b}^{v+3} = \hat{b}^{v+2}$ and $\hat{c}^{v+3} = \hat{c}^{v+2}$.
4. If $\left| l^{v+3} - l^v \right| \leq \eta$ then *stop*, where η is the tolerance value, otherwise, $\hat{\theta}^v = \hat{\theta}^{v+3}$ and move to step 1.

An application of this method to model the Poisson distributed number of deaths is shown by Brouhns, Denuit and Vermunt (2002).

4. Median bilinear model: Least absolute errors

The mean minimizes the sum of squared deviations and the median is the value that minimizes the sum of absolute deviations. The parameters of the median regression are estimated minimizing absolute errors, as follows:

$$\min_{\theta \in \mathbb{R}^{2 \cdot (I+J)}} \sum_{i,j} \left| y_{ij} - a_i - b_j - c_i \cdot d_j \right| \tag{5}$$

The expression (5) can be rewritten as the following minimization problem:

$$\min_{\theta \in \mathbb{R}^{2 \cdot (I+J)}, u \geq 0, v \geq 0} 0.5 \Big(\sum_{i,j} u_{ij} + \sum_{i,j} v_{ij} \Big) \tag{6}$$

subject to

$$y_{ij} - a_i - b_j - c_i \cdot d_j - u_{ij} + v_{ij} = 0.$$

where $u_{ij} = \varepsilon_{ij}$ if $\varepsilon_{ij} > 0$ or 0 otherwise, and $v_{ij} = |\varepsilon_{ij}|$ if $\varepsilon_{ij} < 0$ or 0 otherwise, and $\varepsilon_{ij} = y_{ij} - a_i - b_j - c_i \cdot d_j$. Let us use the following notation $\varepsilon_{ij}(\theta) = y_{ij} - f_{ij}(\theta)$, with $f_{ij}(\theta) = a_i + b_j + c_i \cdot d_j$, to indicate that ε_{ij} depends on the set of parameters θ. Two alternative strategies are adopted to estimate the parameters in (6) based on the conversion of the original nonlinear optimization problem in a sequence of linear problems.

4.1. Strategy A: Linearization of the objective function

Strategy A transforms the nonlinear problem (6) in a sequence of linear problems. Provided that the functions $\varepsilon_{ij}(\theta)$ are continuously derivable in θ, the Lagrangian function may be expressed as $L(s,t,w) = u^{\intercal}(0.5\mathbf{1}_N - s - t) + v^{\intercal}(0.5\mathbf{1}_N + s - w) + \varepsilon(\theta)^{\intercal}s$, where $\mathbf{1}_N^{\intercal}$ is a N-column vector of 1's, $\varepsilon(\theta) = (\varepsilon_{1,1}(\theta), ..., \varepsilon_{ij}(\theta))^{\intercal}$ and s, t and w are the multipliers of Lagrange with t and w are non-negative vectors. Taking partial derivatives with respect to the model parameters θ and the decision variables u and v, we obtain the dual feasibility conditions. The dual version of (6) can be then expressed as,

$$\max_{s \in [-0.5, 0.5]^N} \varepsilon(\theta)^{\intercal}s \qquad s.t \quad J(\theta)^{\intercal}s = 0, \tag{7}$$

where $J(\theta)$ is the vector of first derivatives of $f_{ij}(\theta)$ with respect to θ (El-Attar, Vidyasagar and Dutta, 1979).

4.1.1. Calibration: Affine scale method

Let us consider the locally linearized approximation $\varepsilon(\theta + \Delta) \approx \varepsilon(\theta) - J(\theta) \cdot \Delta$. Koenker and Park (1996) propose to replace $\varepsilon(\theta)$ by the linear approximation $\varepsilon(\theta + \Delta)$ and, then, to apply iteratively the affine scaling method to solve the dual optimization problem (7). Consider the set of initial values $\hat{\theta}^0 = (\hat{a}^0, \hat{b}^0, \hat{c}^0, \hat{d}^0)$. In the iteration step v, parameters are updated as follows:

1. Refine s with Meketon algorithm and estimate Δ which depends on s and $J(\hat{\theta}^v)$, and $\varepsilon(\hat{\theta}^v)$.

2. To ensure that the linearized approximation generates feasible steps, update $\hat{\theta}$ as $\hat{\theta}^{v+1} = \hat{\theta}^v + \lambda\hat{\Delta}$, where $\hat{\Delta}$ is the direction step and λ the length of the step. The length of the step $\lambda \in [0,1]$ is estimated minimizing the primal optimization problem (6) for $\varepsilon(\hat{\theta}^v + \lambda\hat{\Delta})$.

3. If $\sum_{i,j} \left(|\varepsilon_{ij}(\hat{\theta}^{v+1})| - |\varepsilon_{ij}(\hat{\theta}^v)| \right) \leq \eta$ then *stop*, where η is the tolerance value. Otherwise, move to step 4.

4. Project the refined s in the null space of the updated $J(\hat{\theta}^{v+1})$ and rescale to ensure that it is bounded in $[-0.5, 0.5]$, and move to the next iteration.

4.2. Strategy B: Sequence of median linear regressions

Under the strategy B, coefficients in (6) are also estimated by means of an iterative process of a sequence of linear optimization problems. Strategy B draws inspiration from Wilmoth (1993) who replied the method described by Goodman (1979) to the case of minimum least square estimators. Wilmoth (1993) proposed an iterative process to estimate the parameters of the mean bilinear model sequentially by least square techniques. Santolino (2020) adopted this strategy to estimate the parameters of the Lee-Carter quantile mortality model by least absolute techniques. We here describe this strategy for the median bilinear regression. Like the median polish for additive models (Emerson and Hoaglin, 1983), our method relies on the properties of homogeneity and translation invariance satisfied by the median, i.e., for any constant $k \in \mathbb{R}$, the following two equalities are satisfied, $Q_{0.5}(k \cdot y_{ij}) = k \cdot Q_{0.5}(y_{ij})$ and $Q_{0.5}(y_{ij} + k) = Q_{0.5}(y_{ij}) + k$.

Let consider the set of initial values $\hat{\theta}^0 = (\hat{a}^0, \hat{b}^0, \hat{c}^0, \hat{d}^0)$. In the iteration v, parameters are updated as follows:

1. Given $\hat{\theta}^v$, estimate the parameters γ_{a_i} and γ_{b_j} fitting $Q_{0.5}\left(y_{ij}^v\right) = \gamma_{a_i} \cdot \hat{a}_i^v + \gamma_{b_j} \cdot \hat{b}_j^v$, where $y_{ij}^v = y_{ij} - \hat{c}_i^v \cdot \hat{d}_j^v$. Update $\hat{a}_i^{v+1} = \hat{\gamma}_{a_i} \cdot \hat{a}_i^v$ and $\hat{b}_j^{v+1} = \hat{\gamma}_{b_j} \cdot \hat{b}_j^v$, $\hat{c}^{v+1} = \hat{c}^v$ and $\hat{d}^{v+1} = \hat{d}^v$.

2. Given $\hat{\theta}^{v+1}$, estimate the parameter γ_{c_i} fitting $Q_{0.5}\left(y_{ij}^{v+1}\right) = \gamma_{c_i} \cdot \hat{c}_i^{v+1}$, where $y_{ij}^{v+1} = \frac{y_{ij} - \hat{a}_i^{v+1} - \hat{b}_j^{v+1}}{\hat{d}_j^{v+1}}$. Update $\hat{c}_i^{v+2} = \hat{\gamma}_{c_i} \cdot \hat{c}_i^{v+1}$, $\hat{a}^{v+2} = \hat{a}^{v+1}$, $\hat{b}^{v+2} = \hat{b}^{v+1}$ and $\hat{d}^{v+2} = \hat{d}^{v+1}$.

3. Given $\hat{\theta}^{v+2}$, estimate the parameter γ_{d_j} fitting $Q_{0.5}\left(y_{ij}^{v+2}\right) = \gamma_{d_j} \cdot \hat{d}_j^{v+2}$, where $y_{ij}^{v+2} = \frac{y_{ij} - \hat{a}_i^{v+2} - \hat{b}_j^{v+2}}{\hat{c}_i^{v+2}}$. Update $\hat{d}_j^{v+3} = \hat{\gamma}_{d_j} \cdot \hat{d}_j^{v+2}$, $\hat{a}^{v+3} = \hat{a}^{v+2}$, $\hat{b}^{v+3} = \hat{b}^{v+2}$ and $\hat{c}^{v+3} = \hat{c}^{v+2}$.

4. If $\sum_{i,j}\left(|\varepsilon_{ij}(\hat{\theta}^{v+3}| - |\varepsilon_{ij}(\hat{\theta}^{v}|\right) \leq \eta$ then *stop*, where η is the tolerance value. Otherwise, $\hat{\theta}^v = \hat{\theta}^{v+3}$ and move to step 1.

4.2.1. Calibration of a median linear regression

With the application of this strategy, the problem of estimating a median bilinear regression is converted into a problem of estimating iteratively a sequence of three median linear regressions. A median linear regression in matrix notation may be expressed as $Q_{0.5}(Y) = X^{\intercal}\gamma$, where Y is the response vector, γ is the set of parameters to estimate and X is the design matrix. At each step, the following optimization problem has to be resolved:

$$\min_{\gamma, u \geq 0, v \geq 0} 0.5\mathbf{1}_N^{\intercal} u + 0.5\mathbf{1}_N^{\intercal} v \qquad s.t \quad X^{\intercal}\gamma + u - v = Y. \tag{8}$$

Different methods may be applied to estimate the parameters. We briefly describe two estimation methods which are the Mehrotra's Predictor-Corrector method (Portnoy and Koenker, 1997) and the likelihood-based approach (Machado and Silva, 2011; Sánchez et al., 2013).

Mehrotra's Predictor-Corrector method

Alternative algorithms for linear programs with bounded variables may be used to solve (8). A widely used algorithm is the Mehrotra's Predictor-Corrector (MPC) method described in Mehrotra (1992). As the affine scale algorithm, the MPC algorithm belongs to the class of point interior methods. The MPC method is an appropriate algorithm to solve the canonical linear program: $\min\left\{c^{\mathsf{T}}x : Ax = b, \quad x \geq 0\right\}$, where $A \in \mathbb{R}^{mxN}$, y, $b \in \mathbb{R}^m$ and c, x, $s \in \mathbb{R}^N$, and its dual problem, $\max\left\{b^{\mathsf{T}}y : A^{\mathsf{T}}y + s = c, \quad s \geq 0\right\}$. The MPC method finds the joint solution of the primal and dual equations (Salahi, Peng and Terlaky, 2008).

The dual optimization problem of (8) is $\max\left\{y^{\mathsf{T}}s : X^{\mathsf{T}}s = 0, \quad s \in [-0.5, 0.5]^N\right\}$, where s are the multipliers of Lagrange. Setting $a = s + 0.5$, the maximization problem is converted to $\max\left\{y^{\mathsf{T}}a : X^{\mathsf{T}}a = (0.5)X^{\mathsf{T}}\mathbf{1}_N, \quad a \in [0,1]^N\right\}$. Changing the sign of y, it becomes a minimization problem which fits in the setting of the canonical linear program in which the use of MPC method is appropriate.

Maximum likelihood

The likelihood-based approach is based on the asymmetric Laplace distribution to replicate the optimization problem (8). Suppose that the response variable y_l follows an asymmetric Laplace distribution with location parameter $x_l^{\mathsf{T}}\gamma$, scale parameter σ and skewness parameter α, where x_l is the l row of the design matrix, with $l = 1, \ldots, N$. The likelihood function is

$$L(\gamma, \sigma) = \frac{\alpha^N (1-\alpha)^N}{\sigma^N} \exp\left\{ -\sum_{l=1}^{N} \rho_\alpha \left(\frac{y_l - x_l^{\mathsf{T}}\gamma}{\sigma} \right) \right\},$$

where the loss function is defined as $\rho_\alpha(r_l) = r_l(\alpha - I_{r_l})$ for $\alpha \in (0,1)$, I_{r_l} is an indicator function such that $I_{r_l} = 1$ if $r_l < 0$ and zero otherwise. Note that for $\alpha = 0.5$, if σ is considered a nuisance parameter, the maximization of the $L(\gamma, \sigma)$ is equivalent to minimize the objective function (8). Sánchez et al. (2013) describe the steps to obtain the ML estimates based on the expectation-maximization (EM) algorithm.

5. Results

In this section it is compared the performance of mean bilinear models and median bilinear models in presence of extreme values in two different contexts. We illustrate the use of these models with a simulated database and in a real application to the Spanish mortality data. The parameters of the mean bilinear regression model were estimated by least squares (Mean SVD) and also by maximum likelihood (Mean MV). Median bilinear regression models were estimated by the method A and the method B. In the case of the method A, the parameters were estimated by the affine scaling method (Med A-AS). In the case of the method B, we apply the interior point method with the Mehrotra's Predictor-Corrector algorithm (Med B-MP) and the maximum likelihood approach based on the asymmetric Laplace distribution (Med B-MV) to calibrate the model.

All results were calculated in R (R Core Team, 2020). The estimates of the mean bilinear model by maximum likelihood were obtained by means of the *gnm* package (Turner and Firth, 2018) that applies the iterative process in which the linear terms are updated by reweighted least squares (Turner and Firth, 2018; Dutang, 2017). The *nls* function in the *stats* package may be also used to fit a mean bilinear model by a iterative process to minimize least square errors. Median regression models may be estimated by interior point methods with R package *quantreg* (Koenker, 2019), but some implemented functions can only deal with full-rank design matrices. We use the function *rq.fit.fnb* of the package *quantreg* and a version of the function *nlrq* available in Koenker (2020) to calibrate median regression models based on the MPC method and on the Meketon algorithm. Finally, the R package *ALDqr* can be used to estimate median linear models by maximum likelihood (Sánchez et al., 2013). We modify the function *EM.qr* of this package to deal with sparse matrices. The data and code used in the data analysis are publicly available on GitHub (FMBM, 2021)

5.1. Simulation

For illustrative purposes a simulated dataset with extreme values is used for the estimation of median bilinear models. We simulate a database generated by the model (1) in case that the error is normally distributed and there are shocks involving extreme outcomes. Let consider the response variable y_{ij} is generated by $y_{ij} = a_i + b_j + c_i \cdot d_j + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, 0.05)$. The first factor a has 50 levels, $(i \in \{1, \ldots, 50\})$, and the second b has 40 levels, $(j \in \{1, \ldots, 40\})$. The description of coefficients used in the simulation are shown in Table (1).

Table 1. *Descriptive statistics of simulated data.*

	Min.	1st quartile	Median	Mean	3rd quartile	Max.
a_i	1.34	2.95	3.78	3.72	4.85	5.97
b_j	0.00	0.16	0.27	0.31	0.36	0.72
c_i	-2.36	-1.02	-0.42	-0.02	0.2	5.48
d_j	-0.03	-0.02	0.00	0.00	0.01	0.08
y_{ij}	1.32	2.88	4.10	4.04	5.18	7.08

Now we incorporate the extreme outcomes (shocks) to the simulated data. Suppose that the response variable y_{ij} is affected by shocks as follows, $y^s_{ij} = y_{ij} + B \cdot U$, where B is a bernoulli variable that takes 1 with probability p and U is a discrete random variable that takes values $\{-8, -6, 6, 8\}$ with probability 0.25 each one. We consider five different scenarios in relation to the frequency of shocks, $p = (0, 0.01, 0.025, 0.05, 0.1)$, that is, scenario without shocks ($p = 0$), scenario in which the 1% of observations are extreme values ($p = 0.01$), scenario with 2.5% of extreme values ($p = 0.025$), scenario with 5% of extreme values ($p = 0.05$) and scenario with 10% of extreme values ($p = 0.1$).

The mean bilinear and the median bilinear models are fitted to the simulated data in the five scenarios. The sum of squared errors (SSE) and the sum of absolute errors (SAE) are computed for each fitted bilinear model in the scenarios with extreme observations. In order to evaluate the influence of extreme values on estimates, the bilinear models calibrated in the scenarios with extreme observations are also used to compute the statistics of fit for the simulated data without shocks. Results are shown in Table 2.

Table 2. *Statistics of fit of the mean and the median bilinear models.*

		Without shocks	1% shocks		2.5% shocks		5% shocks		10% shocks	
			y^s_{ij}	y_{ij}	y^s_{ij}	y_{ij}	y^s_{ij}	y_{ij}	y^s_{ij}	y_{ij}
Mean SVD	SAE	75.83	338.38	277.75	925.01	670.65	1620.63	1082.47	2803.60	1825.38
	SSE	4.54	730.60	245.61	2686.90	652.01	5620.11	1313.34	10734.41	2905.55
Mean MV	SAE	75.83	338.38	277.75	925.01	670.65	1616.85	1081.19	2803.27	1825.07
	SSE	4.54	730.60	245.61	2686.90	652.01	5600.56	1313.74	10734.41	2905.64
Med A-AS	SAE	75.80	306.88	238.60	922.89	667.86	1504.16	887.85	2707.61	1543.91
	SSE	4.55	852.07	318.87	2687.99	647.77	6102.31	1167.75	12326.93	2613.01
Med B-MP	SAE	74.21	193.53	74.38	487.85	74.79	932.92	75.55	1767.64	77.59
	SSE	4.74	961.50	4.77	3320.13	4.77	6893.69	4.89	13581.87	5.20
Med B-MV	SAE	75.53	195.57	76.91	493.22	82.09	939.44	84.18	1712.79	193.47
	SSE	4.73	956.70	5.01	3300.06	5.64	6872.17	6.01	12848.81	646.84

If we focus on the performance of fitted models in the scenario without shocks (second column of Table 2), as expected, the lowest SSE is observed for the mean bilinear models, and the lowest SAE for the median bilinear models. In fact, this behaviour is repeated in the scenarios with extreme observations when the statistics of fit were computed for the simulated data with the shocks (y^s_{ij}).

However, this conclusion varies when the estimated bilinear models fitted in the scenarios with extreme values are analysed in the scenarios without shocks (y_{ij}). A lower SSE associated to mean bilinear models is not longer observed when the shocks are removed from the simulated data and the statistics of fit are computed again. Now, the fitted median bilinear models show a lower SAE in all scenarios and also a lower SSE in almost all scenarios in comparison with the fitted mean bilinear models. In particular, the performance of the two median bilinear models fitted by method B is clearly better than the performance of the fitted mean bilinear models, and it is also higher than that of the median bilinear model estimated by method A. Comparing between the two median bilinear models fitted by method B, the MPC method seems to provide more stable estimates when the number of extreme values increases. Finally, almost identical outcomes were obtained with the two methods of coefficient estimation for the mean bilinear model.

5.2. Application to mortality data

The second example uses Spanish mortality data to illustrate the application of bilinear models in presence of extreme values. One of the most influential approaches to the stochastic modelling of mortality rates is the parametric nonlinear regression model introduced by Lee and Carter (1992). The Lee-Carter model proposes estimating the conditional mean mortality rate as the nonlinear combination of age and calendar year parameters. Santolino (2020) adopts the Lee-Carter framework to estimate the conditional quantile mortality rate.

The Lee-Carter modelling fits in the setting of the bilinear models defined in (1) in which the main effect of level j (calendar year) is equal to zero, i.e. the response variable (log of the mortality rate) is regressed by the main effect of level i (age) and the interaction between levels i and j. We here estimate the Lee-Carter mean mortality model and the Lee-Carter median mortality model for the Spanish male population. The number of deaths observed, exposures and central mortality rates for the Spanish population by gender were obtained directly from the Human Mortality Database (HMD, 2020). Mortality information is available for ages between 0 and 110, but the number of observations is ineluctably small at the extreme ages and patterns at very advance ages are difficult to observe (Robine et al., 2007). We select ages between 0 and 100 years, which is a common practice in demographics.

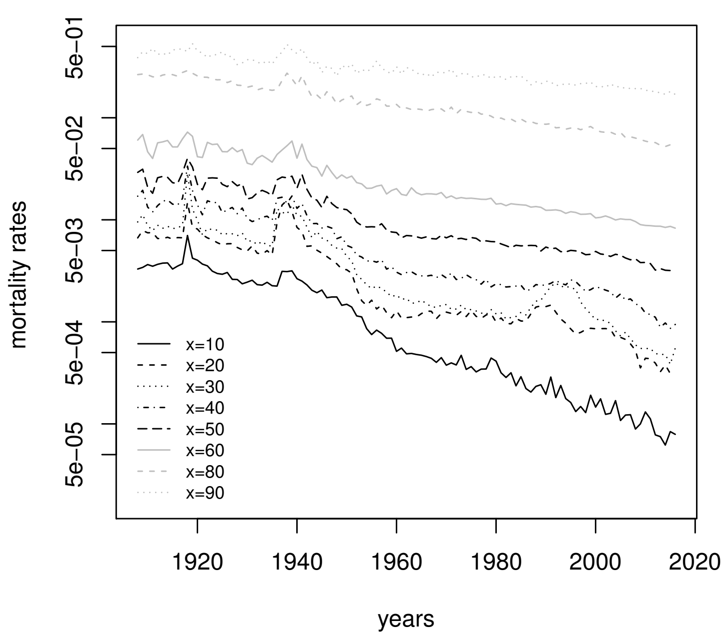

Figure 1. *Mortality rates of Spanish male population at different ages over 1908–2016 period.*

The mortality data cover the observation period between the years 1908 and 2016. Social progress has led to a notable reduction in mortality of the Spanish population through this period. However, there are three spans of time in which the decreasing trend is disrupted, namely, the Spanish flu, the Spanish Civil War and HIV/AIDS. The Spanish flu was a severe influenza pandemic with deadly consequences in 1918 and the following four years (Carreras and Tafunell, 2005). The Spanish Civil War took place between 1936 and 1939. The postwar era formally ended in 1953 with the signing of the US economic agreement (Pact of Madrid). During the war and the first half of the postwar period, poverty and malnutrition affected remarkably the mortality (Jiménez Lucena, 1994). Finally, mortality associated with HIV dramatically increased during the late 80s and 90s, particularly in middle-aged population (CNE, 2011). Figure 1 shows Spanish male mortality rates at different ages in the period 1908–2016, in which these three peaks in the mortality rate are sharply appreciated.

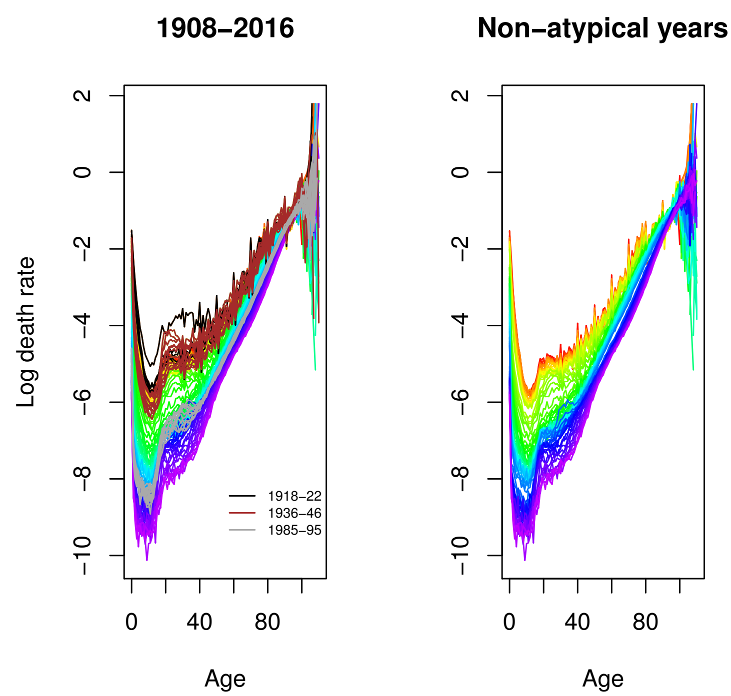

Figure 2. *Mortality rates at different years of Spanish male population.*

The impact of years involving atypically high mortality rate values may be observed in Figure 2. Mortality rates (in log scale) at ages between 0 and 100 are showed for all years in Figure 2 (left). Each line corresponds to the log mortality rates at 0-100 ages in a particular calendar year, and calendar years are differentiated by colors. In case of a continuous reduction in mortality rates over time through the 1908-2016 period, the colored lines should not overlap themselves and they do it. In Figure 2 (right) the years belonging to the time intervals 1918–1922, 1936–1946 and 1985–1995 are not represented. Note that, when these atypical years are removed, the lines seldom overlap themselves.

The mean and median bilinear models are fitted to the Spanish mortality data. The models are calibrated with data involving all calendar years. The measures of goodness-of-fit are computed for the whole sample and when years belonging to the time intervals 1918–1922, 1936–1946 and 1985–1995 are excluded. The sum of squared errors and absolute errors are shown for each fitted bilinear model in Table 3.

Table 3. *Statistics of fit of bilinear models fitted to Spanish male mortality data.*

	All years		Without atypical years	
	SAE	SSE	SAE	SSE
Mean SVD	1279.07	272.76	908.44	172.35
Mean MV	1279.07	272.76	908.44	172.35
Med A-AS	1226.87	300.96	843.84	171.63
Med B-MP	1227.25	301.49	842.90	170.79
Med B-MV	1235.09	288.29	857.19	169.49

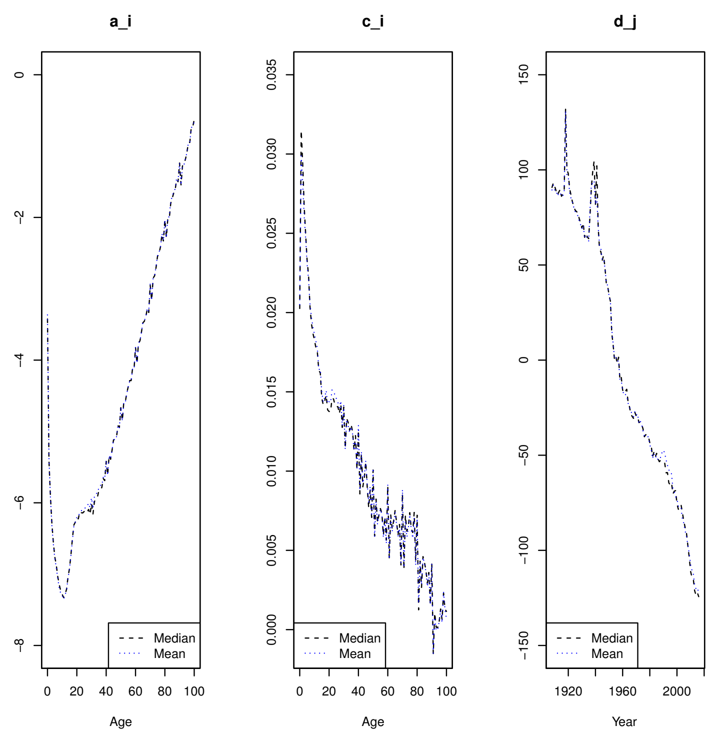

Figure 3. *Coefficient estimates of the Mean MV and Med A-AS models.*

When the goodness-of-fit statistics are computed for the whole sample, the fitted mean bilinear models have lower SSE values and higher SAE values compared to the fitted median bilinear models. Whether or not atypical years are considered in the computation of the statistics of fit, the fitted median bilinear models show lower SSE and SAE values than the fitted mean bilinear models. The performance of the median bilinear models calibrated by the affine scaling method and the MPC method is very similar. Among the fitted median bilinear models, the median bilinear model calibrated by maximum likelihood shows the highest SAE value and the lowest SSE value. Comparing the calibration methods of the mean bilinear model, the same results are obtained with the two methods of coefficient estimation.

A comparison of coefficient estimates of Mean MV and Med A-AS models is provided in Figure 3. Small differences in coefficient estimates are observed for ages in the 20–40 interval and years in time intervals 1936–1946 and 1985–1995.

In mortality applications it is also important to evaluate the prediction power of models. Backtesting is applied to evaluate the prediction accuracy of mean and median models for annual periods up to five years. Alternatively, resampling methods could be used to analyse the prediction power of stochastic mortality models (Atance, Debón and Navarro, 2020). The sum of absolute prediction errors (SAPE) and the sum of squared prediction errors (SSPE) are shown for each bilinear model in Table 4. Median bilinear models show lower SAPE values in all cases and also lower SSPE values in the four-year and five-year forecasting periods (2013–2016 and 2012–2016, respectively).

Table 4. *Backtesting to evaluate prediction power of bilinear models for different periods of forecasting.*

	2016		2015–2016		2014–2016		2013–2016		2012–2016	
	SAPE	SSPE	SAPE	SSPE	SAPE	SSPE	SAPE	SSPE	SAPE	SSPE
Mean SVD	15.14	3.82	28.02	6.90	45.60	11.76	67.07	17.31	82.39	21.04
Mean MV	15.14	3.82	28.02	6.90	45.60	11.76	67.07	17.31	82.39	21.04
Med A-AS	14.52	4.52	27.47	8.89	42.98	13.35	59.46	17.22	73.31	20.83
Med B-MP	14.44	4.51	27.51	8.69	42.89	13.27	59.74	17.30	73.07	20.88
Med B-MV	14.29	4.20	26.99	8.11	42.75	12.75	60.96	17.16	75.00	21.16

6. Conclusions

Conditional mean bilinear regression models have been broadly used in many research fields. In many of the contexts that mean bilinear models are applied, data have extreme observations. It is know that in presence of extreme values the mean may be an inaccurate statistic to reflect the centre of the conditional distribution. In this article we have compared the performance of the mean bilinear model and the median bilinear model in different contexts involving extreme observations.

In the bilinear modelling the multiplicatively interaction structure is specified as a nonlinear term. Alternative methods of parameter estimation for nonlinear regressions are applied. The mean bilinear model is estimated by lest squares and maximum likelihood. The method of parameter estimation for nonlinear median regression involving the linearization of the objective function is compared with the calibration strategy of the median bilinear model in which coefficients are estimated by an iterative process of a sequence of median linear regressions. This second calibration strategy was first used by Santolino (2020) and here it is generalized to the median bilinear model setting.

Mean and median bilinear models are compared in two applications involving extreme values. The first application deals with simulated data with extreme values. The second application is illustrated by means of mortality data of the Spanish population over the 1908–2016 period. During this period, there were a set of years with a particular high mortality (Spanish flu, civil war and HIV/AIDS). Statistics of goodness-of-fit were compared. The fitted median bilinear models showed the lowest sum of absolute errors and the fitted mean bilinear models the lowest sum of square errors. However, when observations with extreme values were removed, the fitted median bilinear models showed the lowest values in the two statistics of goodness-of-fit. This result would confirm that the estimated median is a more appropriate statistic to reflect the centre of the conditional distribution than the estimated mean in these two applications. In the context of COVID-19 using median rather mean approaches when estimating mortality models may be relevant due to the unusual data points arising in 2020 and 2021.

Analysing the two calibration strategies of the median bilinear regression model, we found that the strategy involving the sequence of median linear regressions performed clearly better than the strategy associated to the linearization of the objective function in the application with simulated data and similarly in the application with mortality data.

We conclude that the application of the median bilinear model may be more appropriate than the mean bilinear model in presence of extreme values, whether the centre of the conditional distribution is of interest. Parameters of the median bilinear model may be easily estimated by means of calibrating sequentially median linear regressions. These concluding remarks are relevant in fields such as the stochastic mortality modelling in which researchers have to deal often with data involving extreme observations (wars, pandemics, natural disasters, famines, etc.), and, in general, in any context of application of bilinear models in which the presence of extreme values is frequent.

Acknowledgments

The author thanks Daniel Peña, Montserrat Guillen and participants of the 2020 IREA-UB Annual Seminar for their useful comments. The author acknowledges support received from the Spanish Ministry of Science and Innovation under Grant PID2019-105986GB-C21 and from the Catalan Government under Grant 2020-PANDE-00074.

References

Anderson, J.A. (1984). Regression and ordered categorical variables (with Discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 47, 203–210.

Atance, D., Debón, A., and Navarro, E. (2020) A comparison of forecasting mortality models using resampling methods. *Mathematics*, 8, 1550.

Bassett, G. and Koenker, R. (1978). Theory of least absolute error regression. *Journal of the American Statistical Association*, 73, 618–622.

Black, F., Jensen, M.C. and Scholes, M. (1972). The capital asset pricing model: some empirical tests. *Studies in the theory of capital markets*, New York, Praeger: 79–121.

Brouhns, N., Denuit, M., and Vermunt, J. (2002). A Poisson log-bilinear regression approach to the construction of projected life table. *Insurance: Mathematics and Economics*, 31, 373–393.

Buchinsky, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models a Monte Carlo study. *Journal of Econometrics*, 68, 303–338.

Carreras, A. and Tafunell, X.(2005). *Estadísticas Históricas de España: siglos XIX-XX*. Fundación BBVA, 2a Edic., Bilbao.

CNE (2011). *Área de vigilancia de VIH y conductas de riesgo. Mortalidad por VIH/Sida en España, año 2009. Evolución 1981-2009*. Centro Nacional de Epimediología, Secretaría del Plan Nacional Sobre el Sida, Gobierno de Espaã.

Croux, C., Filzmoser, P., Pison, G. and Rousseeuw, P.J. (2003). Fitting multiplicative models by robust alternating regressions. *Statistics and Computing* , 13, 23–36.

Denis, J.B. and Pázman, A.(1999). Bias of LS estimators in nonlinear regression models with constraints. Part II: Biadditive models. *Applications of Mathematics*, 44, 375–403.

Dutang, C. (2017). Some explanations about the IWLS algorithm to fit generalized linear models. *Technical report*, hal-01577698.

El-Attar, R. A., Vidyasagar, M., and Dutta, S. R. K. (1979). An Algorithm for l_1-norm minimization with application to nonlinear l_1-approximation. *SIAM Journal on Numerical Analysis*, 16, 70–86.

Emerson, J.D. and Hoaglin. D.C. (1983). Analysis of two-way tables by medians. In D. C. Hoaglin, F. Mosteller and J. W. Tukey (Eds.), *Understanding Robust and Exploratory Data Analysis*, 165–210, New York City: John Wiley and Sons.

Erikson, R. and Goldthorpe, J.H. (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press. .Ch. 3.

FMBM (2021). *Fitting Median Bilinear Model*. Available at https://github.com/msantolino/Median-Bilinear-Models. Accessed 8 November 2021.

Gabriel, K.R. (1978). Least squares approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40, 186–196.

Gabriel, K. R. and Odoroff, L. (1984). Resistant lower rank approximation of matrices. In E. Diday, M. Jambu, L. Lebart, J. Pages and R. Tomassone (Eds.), *Data analysis and informatics III*, 23–40, Amsterdam: North-Holland.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74, 537–552.

Goodman, L.A. (1981). Association models and canonical correlation in the analysis of cross- classifications having ordered categories. *Journal of the American Statistical Association*, 76, 320–334.

Hawkins, D.M. (1980). *Identification of outliers*. Monographs on applied probability and statistics, Chapman & Hall.

HMD (2020). *Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)*. Available at www.mortality.org. University of California, Accessed 15 January 2020.

Jiménez Lucena, I. (1994). El tifus exantemático de la posguerra española (1939–1943): el uso de una enfermedad colectiva en la legitimación del 'Nuevo estado'. *Dynamis : Acta Hispanica ad Medicinae Scientiarumque. Historiam Illustrandam*, 14, 185–198.

Justel, A., Peña, D. and Tay, R.S (2001). Detection of outlier patches in autoregressive time series. *Statistica Sinica*, 11, 651–673.

Koenker, R. (2019). *quantreg: Quantile regression*. R package version 5.42.

Koenker, R. (2020). Non linear quantile regression. http://www.econ.uiuc.edu/~roger/research/nlrq/nlrq.html, Accessed 16 February 2021.

Koenker, R. and Park, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, 71, 265–283.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U. S. mortality. *Journal of the American Statistical Association*, 87, 659–671.

Machado, J. and Silva, J. S. (2011). MSS: Stata module to perform heteroskedasticity test for quantile and OLS regressions. Statistical Software Components, Boston College Department of Economics.

Macias, Y. and Santolino, M. (2018). Application of Lee-Carter and Renshaw-Haberman models in life insurance products. *Anales del Instituto de Actuarios Españoles*, 24, 53–78.

Mehrotra, S. (1992). On the implementation of a primal−dual interior point method. *SIAM Journal on Optimization*, 2, 575–601.

Moyano-Silva, P.A., Pérez-Marín, A.M. and Santolino, M. (2020). Estimation of stochastic mortality models for Chile. *Anales del Instituto de Actuarios Españoles*, 4, 225–256.

Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science*, 12, 279–300.

R Core Team, (2020) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Renshaw, A. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality Reduction Factors. *Insurance: Mathematics and Economics*, 38:556–570.

Robine, J.M, Crimmins, E.M, Horiuchi S. and Zeng Yi, Z.(2007). *Human Longevity, Individual Life Duration, and the Growth of the Oldest-Old Population*. International Studies in Population, Springer.

Salahi, M., Peng, J. and Terlaky, T. (2008) On Mehrotra-type predictor-corrector algorithms. *SIAM Journal on Optimization*, 18, 1377–1397.

Sánchez, B., Labros, H. and Labra, V. (2013). Likelihood based inference for quantile regression using the asymmetric Laplace distribution. *Journal of Statistical Computation and Simulation*, 81, 1565–1578

Santolino, M. (2020). The Lee-Carter quantile mortality model. *Scandinavian Actuarial Journal*, 7, 614–633

Turner, H. and Firth, D. (2018). *Generalized nonlinear models in R: an overview of the gnm package*. R package version 1.1-0.

Turner, H., Firth, D. and Kosmidis, I. (2013). Generalized nonlinear models in R. *6th International Conference of the ERCIM WG on Computational and Methodological Statistics, ERCIM*, London.

Van Eeuwijk, Fred A. (1992). Multiplicative models for genotype-environment interaction in plant breeding. *Statistica Applicata*, 4, 393–406.

Van Eeuwijk, Fred A. (1995). Multiplicative interaction in generalized linear models. *Biometrics* , 51, 1017–032.

Wilmoth, J. (1993). *Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change*. Technical Report. Department of Demography. University of California.

Xie, Y. (1992). The log-multiplicative layer effect model for comparing mobility tables. *American Sociological Review*, 57, 380–395.

Yates, F. and Cochran, W.G. (1938). The analysis of groups of experiments. *The Journal of Agricultural Science*, 28, 556–580.